Volume 251, number 1,2, 275–281 FEB 07406 July 1989

Cloning of the pig aminopeptidase N gene

Identification of possible regulatory elements and the exon distribution in relation to the membrane-spanning region

Jørgen Olsen, Hans Sjöström and Ove Norén

Department of Biochemistry C, The Panum Institute, Blegdamsvej 3, University of Copenhagen, Copenhagen, Denmark

Received 5 June 1989

We have isolated four λ-phages covering the complete pig aminopeptidase N/CD13 gene. The sequence of 2.85 kbp encompasses 1.18 kbp of the 5′ upstream region and 1.67 kbp of the structural gene. In the promoter region we find a TATA box and potential binding sites for CTF-1/NF-1 and AP-2. By sequence comparisons we have found three domains showing similarity to promoter regions of the genes encoding human α_1-antitrypsin and human intestinal alkaline phosphatase. The gene sequence includes the first three exons and two introns. It shows that a single exon encodes the cytoplasmic tail, the membrane anchor and the junctional peptide.

Aminopeptidase N; CD13; Gene structure; Regulatory element; Transcription initiation; Exon/intron organization; (Pig)

1. INTRODUCTION

The microvillar membrane of the small intestinal epithelial cell, the enterocyte, is studded with membrane-bound hydrolases that participate in the final hydrolysis of ingested nutrients. Most prominent among these enzymes are the glycosidases and the peptidases (reviews [1,2]). Aminopeptidase N (EC 3.4.11.2) is a 160 kDa metalloprotease belonging to this group of microvillar enzymes. The biosynthesis and intracellular transport of aminopeptidase N has been studied in detail in the enterocyte [3]. Recently, the primary structure of the enzyme was determined by cloning of its cDNA [4] and the gene assigned to chromosome 15q13-qter [5]. Subsequently, it was shown that aminopeptidase N is identical to the myeloid leukemia marker CD13 [6]. A central part of human aminopeptidase N, including a proposed zinc-binding site, shows homology to a bacterial aminopeptidase, suggesting a common evolutionary origin of the two enzymes. However, the homology does not include the cytoplasmic tail, membrane anchor and junctional peptide of human aminopeptidase N [4], indicating that the addition (or conservation) of this segment is unique to the evolution of the membrane-bound enzyme.

Unlike the genes encoding the glycosidases, expression of the aminopeptidase N gene is not solely limited to the small intestine [7]. A high level of expression is also seen in cells lining the kidney proximal tubuli. In the liver and pancreas somewhat lower levels of expression are observed [7]. As a first tep in the understanding of its regulation, we have cloned the pig aminopeptidase N gene and sequenced its promoter. Based on sequence comparisons we have defined three short sequences of potential regulatory importance. Furthermore, the exon/intron arrangement of the first part of the gene shows that the cytoplasmic tail, membrane anchor and junctional peptide are encoded by a single exon supporting the idea that this segment is a separate functional unit of the enzyme.

Correspondence address: J. Olsen, Department of Biochemistry C, Panum Institute, Blegdamsvej 3c, DK-2200 Copenhagen N, Denmark

Published by Elsevier Science Publishers B.V. (Biomedical Division)

2. EXPERIMENTAL

2.1. *Isolation and mapping of the pig aminopeptidase N gene*

Plaque lifts [8] were performed on approx. 1×10^6 plaques from a pig EMBL3 genomic library (Clontech, Palo Alto, CA). Filters were hybridized and washed according to standard procedures [8]. The most stringent wash was performed at 42°C in 15 mM NaCl, 1.5 mM sodium citrate (pH 7.5), 0.1% SDS. The radiolabelled probes used for hybridization were constructed by random priming [9] of a 665 bp 5′ human aminopeptidase N *Pst*I cDNA fragment [4] using [α-^{32}P]dATP. In the primary screening 17 plaques were judged as potentially positive and from these 4 independent hybridizing phages were isolated and designated G2-1, G5-1, G14-1 and G19-1, respectively. Restriction fragments from phage G14-1, together representing 80% of the 15 kbp genomic insert, were subcloned into Bluescript plasmids (Stratagene, San Diego, CA) and mapped with restriction endonucleases. The fragments were ordered with respect to each other by hybridization of the individual fragments to single and double digests of G14-1 DNA. The other isolated phages were digested with the same restriction endonucleases as G14-1 and hybridized with the DNA fragments isolated from this clone.

2.2. *DNA sequencing*

Nested deletions of the two *Eco*RI fragments of 1.10 and 1.75 kbp, originating from G14-1 and subcloned into Bluescript plasmids, were obtained either by the Exo III/mungbean procedure [10] or via the use of internal restriction sites. Preparation of single-stranded DNA and sequencing were performed as in [4]. 65% of the sequence was obtained from both strands. Sequences obtained from one strand only were derived from the sequencing of several independent overlapping clones.

2.3. *Southern blotting*

High molecular mass DNA was extracted from pig spleen as described [11]. 10 μg DNA was digested with restriction endonucleases, electrophoresed through 0.7% agarose at 1.5 V/cm and transferred to nitrocellulose membranes. Filters were hybridized to the labelled fragments under the conditions used for screening the library. Posthybridization washes were performed at 65°C.

2.4. *Primer extension*

RNA was extracted from pig small intestine using the guanidinium-CsCl method [12] and enriched in polyadenylated RNA by chromatography on oligo(dT)-cellulose [8]. 1 pmol of a 17-mer oligonucleotide with the sequence 5′-ATCCCTTGG-CCATGGTG-3′ complementary to nucleotides 42–58 in the pig aminopeptidase N gene was end-labeled with [γ-^{32}P]ATP using T_4 polynucleotide kinase. 0.1 pmol labeled oligonucleotide was mixed with 2 μg poly(A) RNA, heated to 70°C for 10 min and allowed to cool to 42°C. Extensions were performed using 20 U AMV reverse transcriptase (Amersham International) in first strand buffer [13] containing dATP, dCTP, dTTP and dGTP (62.5 μM each). After 45 min the reaction was terminated by extraction with phenol/chloroform, 10 μg *E. coli* tRNA being added before precipitation with ethanol. The precipitated material was dissolved in 4 μl of 98% formamide containing 0.2% (w/v) each of bromophenol blue and xylene cyanol, electrophoresed through a 6% polyacrylamide/urea salt gradient sequencing gel [14] and visualized by autoradiography.

2.5. *S_1 nuclease mapping*

1 pmol of a 1437 bp *Kpn*I/*Eco*RI fragment corresponding to nucleotides −1182 to 255 of the aminopeptidase N gene was dephosphorylated using bacterial alkaline phosphatase and end-labeled with [γ-^{32}P]ATP using T_4 polynucleotide kinase. 0.3 pmol labelled fragment was mixed with 5 μg pig small intestinal poly(A) RNA and 200 μg *E. coli* tRNA in 30 μl of 80% formamide, 40 mM Pipes (pH 6.4), 0.4 M NaCl and 1 mM EDTA. The mixture was heated (80°C, 15 min) and after incubation (3 h, 60°C) 300 μl S_1 buffer [50 mM Na acetate (pH 4.6), 0.28 M NaCl, 4.5 mM $ZnSO_4$, 20 μg/ml denatured herring sperm DNA] containing 200 U/ml S_1 nuclease was added and the mixture further incubated (50 min, room temperature). The S_1 digestion was terminated by addition of 50 μl of 4 M ammonium acetate, 0.1 M EDTA followed by extraction with phenol/chloroform, 20 μg *E. coli* tRNA then being added and nucleic acids precipitated by addition of 380 μl isopropanol. The precipitated material was treated as described for the primer extension experiment.

3. RESULTS

3.1. *Isolation and mapping of the aminopeptidase N gene*

A pig genomic library was screened with a *Pst*I DNA fragment corresponding to the first 665 bp of the human aminopeptidase N cDNA [4]. Four independent hybridizing phages were purified, the DNA being extracted and mapped with restriction endonucleases. One phage (G14-1) was mapped along its entire length (fig.1A). The remaining phages overlapped with G14-1 and were only mapping within the overlapping region. The 1.75 kbp *Eco*RI fragment from clone G14-1 hybridized only to the 5′ specific *Pst*I cDNA fragment. Sequencing of this *Eco*RI fragment confirmed the identity of the gene as it contained a stretch of 525 nucleotides that translated into a 175 amino acid sequence showing extensive homology to the first 180 amino acids of human aminopeptidase N (fig.4). In addition, this *Eco*RI fragment contained 1182 bp of the 5′ upstream region. As indicated in fig.1, G14-1 also hybridized to a 650 bp *Pst*I DNA fragment representing the 3′-end of a rabbit aminopeptidase N cDNA clone. Based on this and the hybridization to phage G5-1 that reaches furtehr 3′ than phage G14-1, we estimate that the entire gene spans approx. 20 kbp. Southern blot analysis of pig genomic DNA (fig.1B,C) confirmed the map predicted from clone G14-1 and did not reveal the existence of additional fragments

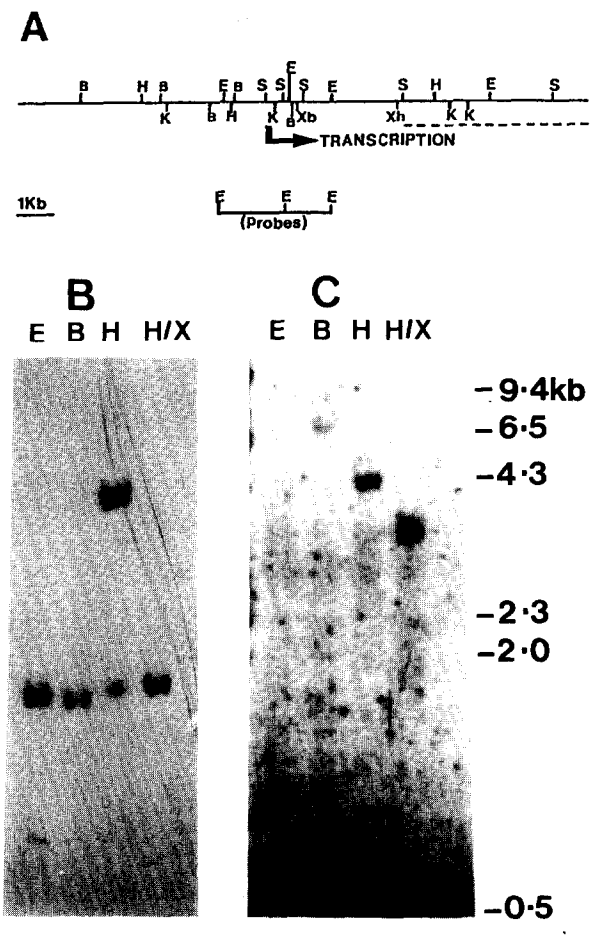

Fig.1. (A) Restriction map of the pig aminopeptidase N gene and 5′ flanking region. Restriction enzymes: *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), *Kpn*I (K), *Sst*I (S), *Xba*I (Xb) and *Xho*I (Xh). The start site of transcription is denoted by an arrow. The dotted line indicates *Sst*I fragments hybridizing to a 650 bp *Pst*I cDNA fragment representing the 3′ part of a rabbit aminopeptidase N clone. (B) Southern blot of pig genomic DNA digested with *Eco*RI (E), *Bam*HI (B), *Hin*dIII (H) and *Hin*dIII/*Xba*I (H/Xb) hybridized with the large *Eco*RI fragment or in (C) the small *Eco*RI fragment indicated in (A). The 3′ *Bam*HI site responsible for the approx. 8 kbp fragment seen in (C) is not found in the genomic insert of phage G14-1 (but is found within phage G5-1) and is therefore not shown.

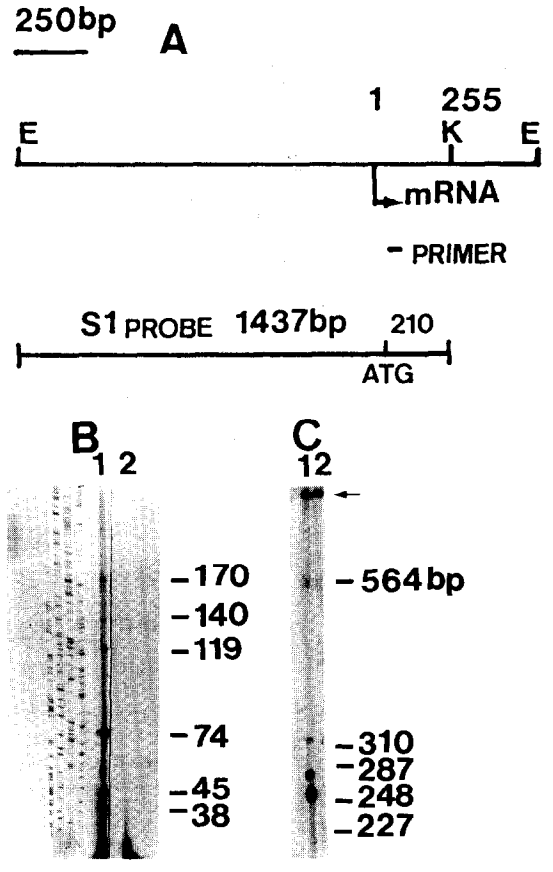

Fig.2. Mapping of the transcriptional start sites of the aminopeptidase N gene. (A) Schematic representation of the 1.75 kb *Eco*RI fragment from phage G14-1. *Eco*RI (E) and *Kpn*I (K) sites are shown. (B) (Lane 1) Primer extension of 2 μg pig small intestinal poly(A) RNA; (lane 2) primer alone. The position of the primer is shown in (A). The primer extension products are electrophoresed along a sequence reaction of a *Pst*I fragment (nucleotides −55 to 543) subcloned into bluescript SK^+ and primed with the oligomer used for primer extension. The numbers indicate the distance to the AUG codon in the mRNA. (C) S_1 mapping of pig small intestinal poly(A) RNA (lane 1) and *E. coli* tRNA (lane 2). Arrow indicates the intact probe. The probe used is shown in (A).

hybridizing to aminopeptidase N coding sequences.

3.2. *Initiation of transcription*

The 5′-end of aminopeptidase N mRNA was mapped by primer extension of mRNA extracted from pig small intestine (fig.2B). Initiation sites are observed 38, 45, 74, 119, 140 and 170 bp upstream from the initiator ATG triplet. The signals 38, 45 and 74 bp upstream from the ATG triplet are the most intense. With S_1 nuclease mapping (fig.3C) three protected fragments of 255, 285 and 330 bp respectively, are observed. This confirms the presence of the initiation sites 45, 74 and 119 bp upstream from the ATG triplet. The site defined by primer extension 38 bp upstream from

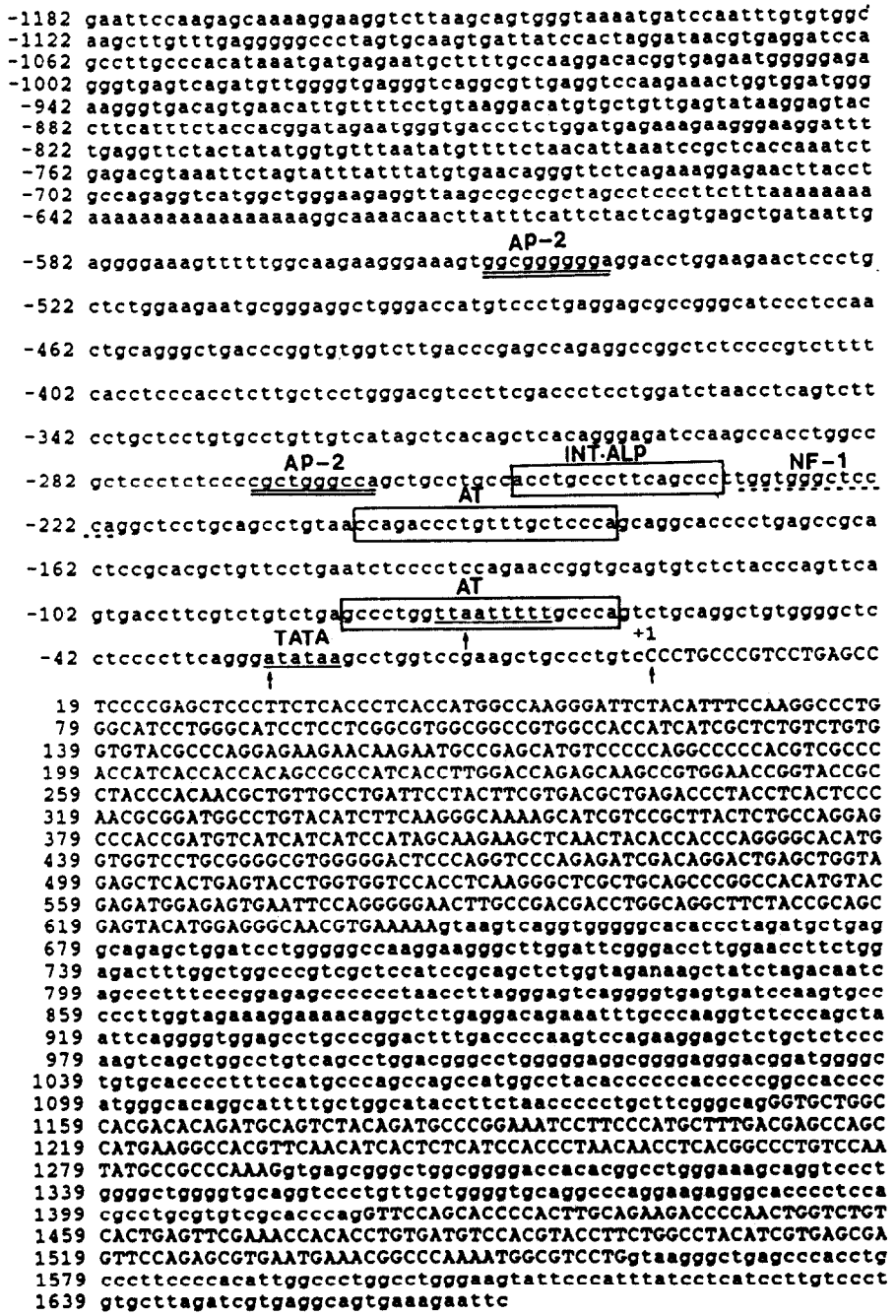

Fig.3. Sequences flanking the transcriptional start sites of the pig aminopeptidase N gene. 5′ upstream region as well as introns are in lower-case letters while exon sequences are in capitals. Sequences of potential importance are indicated as follows: the TATA box and TATA-like sequences are underlined; regions of homology to the promotors of human α_1-antitrypsin [22] and human intestinal alkaline phosphatase [23] are boxed designated AT and INT·ALP respectively. The dashed line indicates a potential binding site for NF-1/CTF-1 [16]. Two potential binding sites for the factor AP-2 [19] are double underlined.

the ATG triplet is not seen in the S_1 nuclease mapping experiment. It is not known whether the primer extension signal seen at this position is an artefact or if the S_1 nuclease fails to distinguish between these sites under the mapping conditions used. In the S_1 mapping experiment there is no evidence – not even after prolonged exposure of the autoradiograph – of mRNA reaching further upstream than the messenger initiated 119 bp upstream from the ATG triplet. As the initiation site 45 bp upstream from the ATG triplet is the first site defined by both primer extension and S_1 nuclease mapping, it is used to define position 1 in the gene sequence.

3.3. *Sequences of the aminopeptidase N gene*

The two *Eco*RI fragments of 1.10 and 1.75 kbp were sequenced (fig.3). The sequence information obtained covered 1.18 kbp of the upstream region and 1.67 kbp of the coding part of the gene.

3.4. *Sequence analysis of the 5′ upstream region*

The 5′ upstream region was searched for putative binding sites for known transacting factors [15]. At position −29 there is a 'perfect' TATA box – ATATAA – and at −76 to −68 a TATA-like sequence – TTAATTTTT – is observed. From position −233 to −221 a palindromic sequence that matches the consensus $TGGN_7CCA$ is found. This sequence is known to interact with a family of transacting factors to which NF-1/CTF-1 and RPF-1 [16–18] belong. Two potential binding sites for the transacting factor AP-2 [19] are found (in reverse) at position −270 to −261 and −552 to −543 in the aminopeptidase N gene promoter region.

```
Porcine  amp.N        -83 GCCCTGGTTAATTTTTGCCCA
Human α1-AT           -80 ACCTTGGTTAATATTCACCAG

Porcine amp.N        -202 CCAGACCCTGTTTGCTCCCA
Human α1-AT          -114 TTAGCCCCTGTTTGCTCCTC

Porcine amp.N        -250 ACCTGCCCTTCAGCCC
human int.alp.       -192 CCCTTCCCTTCAGCAA
```

Fig.5. Similarity to other promotors. The sequences in the aminopeptidase N promotor showing similarity to regions in the human α_1-antitrypsin promotor [22] and the human alkaline phosphatase promotor [23] are aligned with the relevant sequences of these promotors.

3.5. *Sequence analysis of aminopeptidase N gene coding region*

The 1.67 kbp coding region contains the first three exons with the two intervening introns (figs 3,4). Overall there is high homology (79%) between the encoded 294 amino acid sequence of the pig enzyme and the first 301 amino acids of the human enzyme although gaps must be introduced to obtain an optimal alignment (fig.4). The exon/intron junctions are AA/GT, AG/GT, TG/GT for the first, second and third junctions, respectively. The intron/exon junctions are CAG/G for both junctions. The junctions are in agreement with those of other vertebrate genes [20].

```
  1 MAKGFYISKALGILGILLGVAAVATIIALSVVYAQEKNKNAEHVPQAPTSPTITTT----
    --------+S----------C----------+S-------NSS-V-S-T-SASA-TNPA

 57 AAITLDQSKPWNRYRLPTTLLPDSYFVTLRPYLTPNADGLYIFKGKSIVRLLCQEPTDVI
    S-T------A-------N--K----R----------DR-------S-T--FT-K-A----

117 IIHSKKLNYTT-QGHMVVLRGVGDSQVPEIDRTELVELTEYLVVHLKGSLQPGHMYEMES
    ----------LS---R---L---G--P-D--K-----P----------VKDSQ---D-

                    exon 1 /exon 2
176 EFQGELADDLAGFYRSEYMEGNVKKVLATTQM--QSTDARKSFPCFDEPAMKATFNITLI
    --E-------------------R--E-----QM-AA----------------E------

          exon 2/exon 3
234 HPNNLTALSNMPPKGSSTPLAEDPNWSVTEFETTPVMSTYLLAYIVSEFQSVNETAQNGV
    --KD-------L---P----P-----N----H---K-------F-----DY-EKQ-S---

294 L
    -
```

Fig.4. Amino acid sequence encoded by the first three exons of the pig aminopeptidase N gene. The predicted amino acid sequence is aligned with the corresponding part of the human enzyme. Identical residues are indicated with a line and the exon junctions indicated above the sequence. The membrane anchor is boxed and the serine/threonine-rich stretch is underlined.

4. DISCUSSION

4.1. *Promoter elements*

The TATA sequence at position −29 is probably responsible for directing transcription from position 1. Of the two potential AP-2 binding sites the first (position −270 to −261) is degenerate at position two (a guanine) compared to the consensus, whereas the other (position −552 to −543) matches the consensus 5′ T/C C C/G CC A/C N G/C C/G G/C 3′ [19]. AP-2 is thought to be able to induce transcription both by stimulation with phorbol esters through a protein kinase C-dependent pathway as well as via forskolin by a cAMP-dependent pathway. Intestinal mucosal explants cultured in the presence of forskolin showed a reduced surface expression of the mature form of aminopeptidase N. However, the total pool of newly synthesised enzyme was unaffected [21]. The transcription rate of the aminopeptidase N gene in cultured enterocytes might already be maximally stimulated by cAMP without the addition of forskolin, thus explaining the absence of an effect on the biosynthetic rate. In contrast, the finding could indicate that the binding sites for AP-2 are not of great importance for the expression of the aminopeptidase N gene. The latter interpretation is further supported by the fact that AP-2 binding sites are not found in this region of the human gene (Shapiro, L.H. and Look, A.T., personal communication).

4.2. *Similarity to other promoters*

The aminopeptidase N promoter was compared with other promoters of genes known to be expressed in small intestine and liver. Fig.5 shows an alignment of three sequences in the aminopeptidase N promoter that display similarities to regions in the promoters of the genes encoding human α_1-antitrypsin [22] and human intestinal alkaline phosphatase [23]. The sequences are also found in the human aminopeptidase N gene (Shapiro, L.H. and Look, A.T., personal communication) increasing the significance of the similarity. The sequences from the human α_1-antitrypsin promoter have been demonstrated to interact with two proteins purified from rat liver [22]. As the aminopeptidase N gene is efficiently expressed in liver it is highly probable that these sequences are of importance for the expression of the gene in this tissue. It is not yet known if the sequence in the intestinal alkaline phosphatase promoter interacts with DNA-binding proteins. It should be noted that this sequence is also found in the gene encoding the placental form of alkaline phosphatase [24]. As aminopeptidase N is found in the placenta, a factor common to small intestine and placental tissue could be responsible for the expression of the gene in these tissues. This, however, requires additional regulatory mechanisms for the alkaline phosphatase genes to explain their tissue-specific expression.

The aminopeptidase N activity found in different tissues varies over three orders of magnitude [7], indicating powerful tissue-specific regulation of aminopeptidase N gene expression. This is also reflected in the structure of the promoter region which does not bear the characteristics of constitutively expressed genes (the so-called housekeeping genes), as it has a distinct TATA box, a moderate G + C content and a low CpG to GpC ratio. In addition, it lacks an SP1 binding site often seen in promoters of housekeeping genes.

4.3. *Coding sequences – implications for the evolution of the aminopeptidase N gene*

Exon 1 encodes the cytoplasmic tail, membrane anchor and so-called junctional peptide of the enzyme. The second exon/intron junction appears just before the start of the region showing homology to a bacterial aminopeptidase [4]. This supports our previous hypothesis that the genes encoding the bacterial and mammalian aminopeptidase N evolved from a common ancestral gene. During the course of evolution, a segment encoding the cytoplasmic tail, membrane anchor and junctional peptide has been added (or preserved) to the mammalian enzyme making it an intestinal membrane protein, enabling it to carry out effectively surface hydrolysis.

Acknowledgements: Ms Liselotte Laustsen and Jette Møller are thanked for excellent technical assistance and Dr Martin Spiess (Dept of Biochemistry, Biozentrum, University of Basel) for providing oligonucleotides. We are grateful to Drs L.H. Shapiro and T. Look (St Jude Children Res. Hospital, Memphis, TN) for providing unpublished information on the human aminopeptidase N gene. The work was supported by the Danish Medical Research Council (12-8115, 12-7331), Michaelsen Foundation and Lundbeck Foundation. The project was part of a program under the Biomembrane Research Center, Aarhus University, Denmark.

REFERENCES

[1] Norén, O., Sjöström, H., Danielsen, E.M., Cowell, G.M. and Skovbjerg, H. (1986) in: Molecular and Cellular Basis of Digestion (Desnuelle, P. et al. eds) pp.335–365, Elsevier, Amsterdam.

[2] Semenza, G. (1986) Annu. Rev. Cell Biol. 2, 255–313.

[3] Danielsen, E.M., Cowell, G.M., Norén, O. and Sjöström, H. (1987) in: Mammalian Ectoenzymes (Kenny, A.J. and Turner, A.J. eds) pp.47–85, Elsevier, Amsterdam.

[4] Olsen, J., Cowell, G.M., Kønigshøfer, E., Danielsen, E.M., Møller, J., Laustsen, L., Hansen, O.C., Welinder, K.G., Engberg, J., Hunziker, W., Spiess, M., Sjöström, H. and Norén, O. (1988) FEBS Lett. 238, 307–314.

[5] Kruse, T.A., Bolund, L., Grzeschik, K.H., Ropers, H.H., Olsen, J., Sjöström, H. and Norén, O. (1988) FEBS Lett. 239, 305–308.

[6] Look, A.T., Ashmun, R.A., Shapiro, L.H. and Peiper, S.C. (1989) J. Clin. Invest., in press.

[7] Rehfeld, N., Peters, J.E., Giesecke, H., Beier, L. and Haschen, R.J. (1967) Acta Biol. Med. Germ. 19, 809–818.

[8] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

[9] Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.

[10] Heinkoff, S. (1984) Gene 28, 351–359.

[11] Herrmann, B.G. and Frischauf, A.M. (1987) Methods Enzymol. 152, 180–183.

[12] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.

[13] Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269.

[14] Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963–3965.

[15] Jones, N.C., Rigby, P.W.J. and Ziff, E.B. (1988) Genes Devel. 2, 267–281.

[16] Gronostajski, R.M. (1987) Nucleic Acids Res. 15, 5547–5559.

[17] Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. and Tjian, R. (1987) Cell 48, 79–89.

[18] Gil, G., Osborne, T.F., Goldstein, J.L. and Brown, M.S. (1988) J. Biol. Chem. 263, 19009–19019.

[19] Imagava, M., Chiu, R. and Karin, M. (1987) Cell 51, 251–260.

[20] Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Annu. Rev. Biochem. 55, 1119–1150.

[21] Danielsen, E.M., Hansen, G.H. and Cowell, G.M. (1987) Eur. J. Cell Biol. 44, 273–277.

[22] Hardon, E.M., Frain, M., Paonessa, G. and Cortese, R. (1988) EMBO J. 7, 1711–1719.

[23] Henthorn, P.S., Raducha, M., Kadesch, T., Weiss, M.J. and Harris, H. (1988) J. Biol. Chem. 263, 12011–12019.

[24] Knoll, B.J., Rothblum, K.N. and Longley, M. (1988) J. Biol. Chem. 263, 12020–12027.